

# The Impact of Data Sharing on Data Quality

Eric D. Whitman, M.D., Mark E. Frisse, M.D., Michael G. Kahn, M.D., Ph.D.,  
Washington University, St. Louis, MO

This study was undertaken to better understand the relationship between private (developed by the clinician to address specific areas of interest), institutional, and regional health care databases. Potential data storage inefficiencies in our private database were correctable by sharing data between these entities.

## METHODS

The Vascular Access Service (VAS) is a clinical specialty service at Washington University Medical School that specializes in the placement, maintenance, and removal of venous access devices. The VAS maintains a database (VASD) with 1548 records, each representing a separate device. The VASD stores information about device implantation, complications (such as infection and thrombosis), and removal. Presently, the VASD does not exchange data with the enterprise IS. The latter database contains demographic, laboratory and radiologic information valuable to the validity of the VASD.

### Data Interaction Scenarios

The hospital IS was queried to investigate the impact of data sharing upon VASD data quality in two separate situations:

**Scenario #1.** If the VASD records the placement of a long term device, but not a removal date, does that necessarily mean that the device is still in place (measured in days as dwell time)? One possible reason for "removal", patient death, could be obtained from the hospital IS information, if the patient died while an inpatient.

**Scenario #2.** If the VASD records the implantation and removal of a device, without infectious problems when removed, does this mean that the device was never infected? Often the only clue that the device might be infected is a positive blood culture (septicemia), available from the hospital IS.

For both scenarios, key clinical statistics were calculated before and after the VASD was updated using data stored in the enterprise IS. Differences in data calculations affected by the updating were compared using the t-statistic (Table 1).

**Table 1**

Total VASD records:	1548
<b>Scenario #1</b>	
Records meeting criteria	648 (42%)
Dwell time	418±10 days
Expired Patients	102 (15.7%)
Updated dwell time	366±10 days *
Total Variance	34,049 days
Mean variance	334 days/device
Range of variance	12-910 days
<b>Scenario #2</b>	
Records meeting criteria	237 (15%)
Number with potentially infected devices	45 (19%)

\*  $p < 0.0001$ , compared to pre-update values

## DISCUSSION

A private database, like the VASD, contains information that is highly specialized and oriented towards a specific clinical concept, such as a disease, medication, or surgical procedure. The advantage of the VASD is its highly specialized nature, which permits immediate ad hoc queries about these devices that are otherwise impossible. Unfortunately, our current inability to exchange data among the VASD, enterprise-wide, and regional databases creates significant potential data errors and inconsistencies.

The VASD lacks an efficient means for data validation and update. If the clinical team responsible for the VASD isn't informed of the event, or doesn't perform a manual data update, data errors accumulate, leading to mistakes in calculating the dwell time (15.7% correction rate based on institution IS updates) and the incidence of infection (19% correction rate). If the VASD shared data with a regional IS, outpatient events could also be updated. These problems adversely impact the quality and cost-effectiveness of our clinical care. For example, the risk of infection may alter the type (and cost) of device chosen in a given clinical situation.

Private, hospital, and regional IS coexist within our health care system, but the nature and extent of their optimal interaction is unclear. These data suggest that communication between personal and enterprise-wide databases could lead to better informed, more cost-effective clinical management decisions.